

Identification of Disease Genes

This exercise focuses on the identification of a disease gene using NCBI's human genome assembly. The reference human genome assembly along with integrated maps, literature, and expression information comprises a powerful discovery system for exploring candidate human disease genes.

Problem: A laboratory has generated an EST library from a hemochromatosis patient and wants to identify the gene(s) causing the phenotype.

We will follow these steps to solve the problem:

1. Compare an EST to the human genome (using BLAST).
2. Identify the gene(s) aligning with the EST and download their sequences (using MapViewer).
3. Identify whether the EST contains any known SNPs (using dbSNP).
4. Determine whether a mutant form of the gene causes a phenotype (using OMIM).

Problem 1:

A laboratory has generated an EST library from a hemochromatosis patient and wants to identify the gene(s) causing the phenotype.

Outline:

We will follow these steps to solve the problem:

1. Compare the EST from a hemochromatosis patient to the human genome (using BLAST).
2. Identify the gene(s) aligning with the ESTs and download their sequences (using Map Viewer).
3. Identify whether the EST contains any known nucleotide variations (single nucleotide polymorphisms) (using dbSNP).
4. Determine whether a mutant form of the gene is known to cause a phenotype (using OMIM).

Step 1. Compare ESTs to the human genome (using BLAST):

One way to identify the genes expressing the EST is to compare the EST sequence with the human genome assembly and the genes annotated on it. To

access the specialized BLAST page for searching against the human genome assembly, click on

[BLAST \(human genome\)](#)

Copy the EST sequence provided below. Paste it in the query box of the BLAST page and select the "genome (reference only)" database. Start the search by clicking on the "Begin Search" button.

Query EST Sequence:

```
TGCCTCCTTTGGTGAAGGTGACACATCATGTGACCTCTTCAGTGACCACTCT
ACGGTGTGCGGGCCTTGAACCTACTACCCCCAGAACATCACCATGAAGTGGCT
GAAGGATAAGCAGCCAATGGATGCCAAGGAGTTTGAACCTAAAGACGTATT
GCCCAATGGGGATGGGACCTACCAGGGCTGGATAACCTTGGCTGTACCCC
CTGGGGAAGAGCAGAGATATACGTACCAGGTGGAGCACCCAGGCCTGGAT
CAGCCCCTCATTGTGATCTGGG
```

Name the chromosome and the contig that we get as a BLAST hit. Is the EST sequence 100% identical to the genomic sequence? Note the nucleotide difference and position between the two sequences.

Step 2. Identify the gene(s) expressing the ESTs and download their sequences:

To visualize the BLAST hit on the genome using Map Viewer, click on the "Genome View" button at the top of the results page, then on the Map element "NT_007592". Currently, 4 maps should be displayed (Contig, Model, RNA and Gene_seq). Zoom out 2 or 4 times by clicking on right most contig map and selecting the appropriate option.

The BLAST hit, indicated by the red bar, is in the region of one of the exons of the HFE gene annotated on the human genome. Make the Gene_seq map a master map by clicking on the arrow at the top of the map. Display the entire HFE gene sequence by clicking on the "dl" link and then on "Display". Copy the sequence and paste it in the area provided below. We will use it later to obtain the exon-intron structure. You can adjust the nucleotide locations to download the upstream or downstream sequence by using the "adjust by" and "Change Region/Strand" option.

Step 3. Determine whether the ESTs contain known SNPs:

Go back to the Map Viewer report. Click on the Maps and Options link. Remove all the maps except the Gene_seq map by selecting the map under the Maps Displayed menu and clicking on Remove. Now add the variation map from the Available maps menu (by selecting the map and clicking on Add). Make the Variation map as the master map by selecting it and clicking the Make

Master/Move to Bottom option. Then click on Apply. Now two maps are displayed, Variation (it's the rightmost and master map) and Gene_seq. The master map provides detailed information for the map features, in this case SNPs. ". (The Mini-Course Map Viewer Quick Start describes the usage of the Map Viewer in detail.) Zoom in on the blast hit area (red bar). There are two SNPs in the area, one of them is rs1800562. Click on the link for the SNP. There is an A/G SNP is at the nucleotide position 26033141 on the contig NT_007592 as mentioned under Fasta sequence and Integrated maps. Is this the same nucleotide variation found in the BLAST result in Step 1? Please note that the SNP results in the Cysteine 282 Tyrosine mutation for the longest protein (expressed by the mRNA NM_000410) as reported under GeneView.

Step 4. Determine whether the mutant HFE gene causes a phenotype:

Go back to the Map Viewer report and add the Phenotype map as the master map using the Maps and Options menu. Click on the HFE link that leads to the hemochromatosis record (235200) in the OMIM database. The record details how mutations in the HFE gene are associated with a phenotype, hemochromatosis. Click on the Allelic Variant "View list" to get information about mutant proteins from patients. Is Cys282Tyr variant mentioned in the list? Which phenotype does it cause?

Summary:

This exercise describes steps to identify the gene expressing an EST obtained from a hemochromatosis patient, download the gene sequence, identify known SNPs in the gene and find SNP-associated phenotypes.

Step 1: The query EST sequence was found to align contig NT_007592.15 on chromosome 6 with one nucleotide difference (G to A with respect to the nucleotide 26033141 on the contig).

Step 2: The query EST was found to be expressed by the HFE gene.

Step 3: The query EST sequence contains a known SNP (G/A with respect to nucleotide 16951392 on contig NT_007592.15).

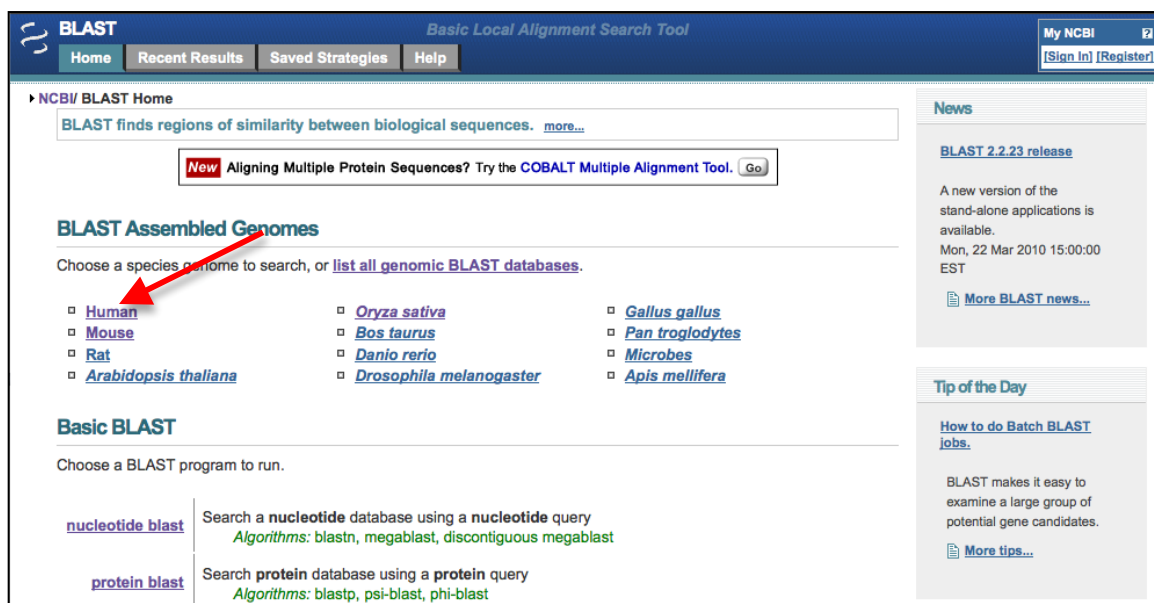
Step 4: Mutations in the HFE gene are associated with hemochromatosis.

Screen images


Step 1: Compare EST against the human genome.



The screenshot shows the NCBI homepage. At the top, there is a navigation bar with 'NCBI', 'Resources', and 'How To'. Below this is a search bar with a dropdown menu set to 'All Databases'. The main content area is divided into several sections: 'Resources' on the left with a list of links like 'NCBI Home', 'All Resources (A-Z)', 'Literature', etc.; 'Welcome to NCBI' in the center with a brief description and links to 'More about the NCBI'; 'PubMed Central' with a featured article; 'How To...' with a list of tasks like 'Obtain the full text of an article'; 'Popular Resources' on the right with links to 'PubMed', 'PubMed Central', 'Bookshelf', 'BLAST', etc.; and 'NCBI News' with recent updates. A red arrow points to the 'BLAST' link in the 'Popular Resources' section.



The screenshot shows the BLAST homepage. At the top, there is a navigation bar with 'BLAST', 'Basic Local Alignment Search Tool', and 'My NCBI'. Below this is a 'Home' button. The main content area is divided into several sections: 'NCBI/ BLAST Home' with a description of BLAST; 'BLAST Assembled Genomes' with a list of species to search, including 'Human', 'Mouse', 'Rat', 'Arabidopsis thaliana', 'Oryza sativa', 'Bos taurus', 'Danio rerio', 'Drosophila melanogaster', 'Gallus gallus', 'Pan troglodytes', 'Microbes', and 'Apis mellifera'; 'Basic BLAST' with a list of BLAST programs like 'nucleotide blast' and 'protein blast'; and 'News' with a recent update about 'BLAST 2.2.23 release'. A red arrow points to the 'Human' link in the 'BLAST Assembled Genomes' section.



[NCBI Home](#) [Genomic Biology](#) [Human Genome Resources](#) [BLAST](#)

Search

Map Viewer

Go

Clear

BLAST
[Overview](#)
[FAQs](#)
[News](#)
[Manual](#)
[References](#)
[Retrieve results](#)

Genome Project

BLAST Human Sequences.

An alternate informative description.

☒ Enter an accession, gi, or a sequence in FASTA format:

```
TGCCTCCTTTGGTGAAGGTGACACATCATGTGACCTCTTCAGTGACCACTCTACGGTGTCTGGGCCTT
GAACTACTACCCCAAGAACATCACCATGAAGTGGCTGAAGGATAAGCAGCCAATGGATGCCAAGGAG
TTCGAACCTAAAGACGTATTGCCCAATGGGGATGGGACCTACCAGGGCTGGATAACCTTGGCTGTAC
CCCTGGGGAAGAGCAGAGATATACGTACCAGGTGGAGCACCAGGCTGGATCAGCCCTCATTGT
GATCTGGG
```

☐ Or, choose a file to upload

Browse...

Set subsequence: (optional)

From: To:

Database:

genome (reference only)

259 sequences

Program:

megaBLAST: Compare highly related nucleotide sequences

Optional parameters

Expect	Filter	Descriptions	Alignments
0.01	default	100	100

Advanced options:

Begin Search

Clear Input

☐ auto-check for results

Get the URL with preset values ?

Get URL

BLAST Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

► **NCBI/ BLAST/ Format Request**

Query Nucleotide Sequence (276 letters)
 Database ref_contig
 Job title Nucleotide Sequence (276 letters)
 Request ID WCSRENPU01S **View report** ☐ Show results in a new window

Format

Show Alignment as HTML ☐ Advanced View ☐ Use old BLAST report format [Reset form to defaults](#)

Alignment View Pairwise

Display ☒ Graphical Overview ☒ Linkout ☒ Sequence Retrieval ☐ NCBI-gi ☐ CDS feature

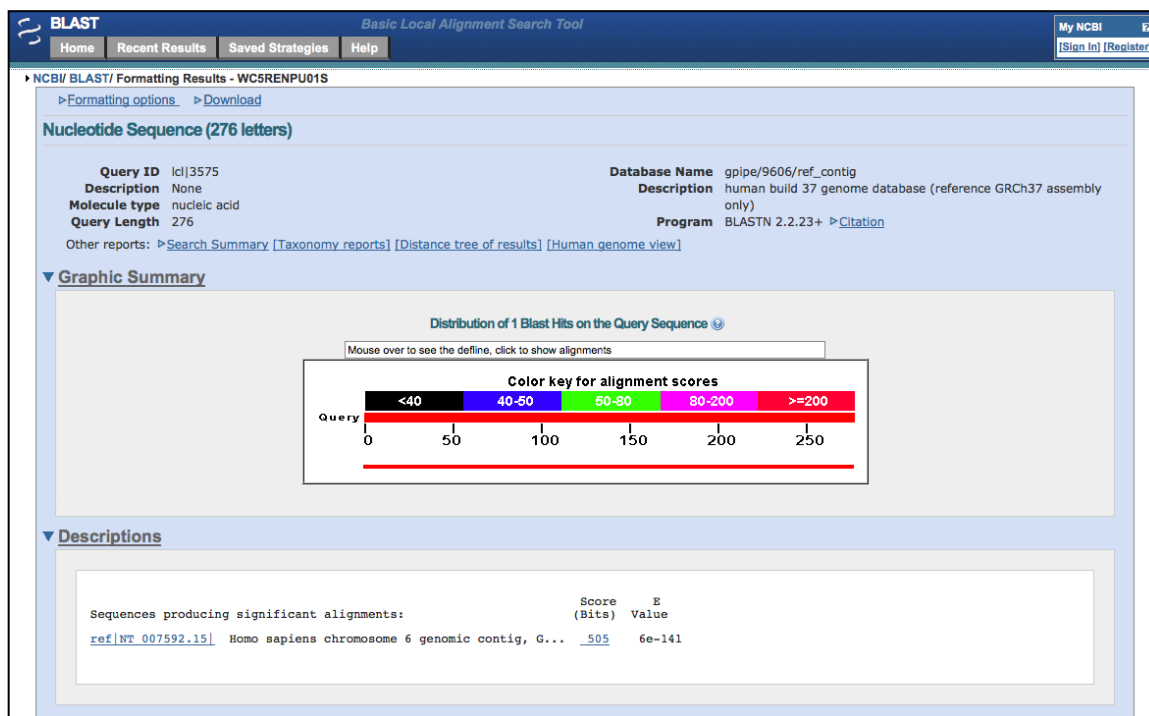
Masking Character: Lower Case Color: Grey

Limit results Descriptions: 100 Graphical overview: 100 Alignments: 100

Organism Type common name, binomial, taxid, or group name. Only 20 top taxa will be shown.

Entrez query:

Expect Min: Expect Max:



Alignments

☐ Select All [Get selected sequences](#) [Distance tree of results](#)

> [ref|NT_007592.15|](#) **D** Homo sapiens chromosome 6 genomic contig, GRCh37 reference primary assembly
Length=58720166

Features in this part of subject sequence:
[hemochromatosis protein isoform 1 precursor](#)
[hemochromatosis protein isoform 6 precursor](#)

Score = 505 bits (273), Expect = 6e-141
 Identities = 275/276 (99%), Gaps = 0/276 (0%)
 Strand=Plus/Plus

Query 1	TGCCTCCTTTGGTGAAGGTGACACATCATGTGACCTCTTCAGTGACCACTCTACGGTGTC	60
Sbjct 26032913	TGCCTCCTTTGGTGAAGGTGACACATCATGTGACCTCTTCAGTGACCACTCTACGGTGTC	26032972
Query 61	GGGCCTTGAACCTACTACCCCAAGACATCACCATGAAGTGGCTGAAGGATAAGCAGCCAA	120
Sbjct 26032973	GGGCCTTGAACCTACTACCCCAAGACATCACCATGAAGTGGCTGAAGGATAAGCAGCCAA	26033032
Query 121	TGGATGCCAAGGAGTTCGAACCTAAAGACGTATTGCCCAATGGGGATGGGACCTACCAGG	180
Sbjct 26033033	TGGATGCCAAGGAGTTCGAACCTAAAGACGTATTGCCCAATGGGGATGGGACCTACCAGG	26033092
Query 181	GCTGGATAACCTTGGCTGTACCCCTGGGGAAGAGCAAGATATACGTACCAGGTGGAGC	240
Sbjct 26033093	GCTGGATAACCTTGGCTGTACCCCTGGGGAAGAGCAAGATATACGTACCAGGTGGAGC	26033152
Query 241	ACCCAGGCCTGGATCAGCCCTCATTGTGATCTGGG	276
Sbjct 26033153	ACCCAGGCCTGGATCAGCCCTCATTGTGATCTGGG	26033188

Result: The EST sequence is aligned to the contig NT_007592.15 on chromosome 6 with one nucleotide difference (G to A with respect to the nucleotide 26033141 on the contig).

Step 2: Identify the gene(s) expressing the EST and download their sequences

Query ID lcl|3575
Description None
Molecule type nucleic acid
Query Length 276

Database Name gpipe/9606/ref_contig
Description human build 37 genome data (only)
Program BLASTN 2.2.23+ [Citation](#)

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [Human genome view](#)

Graphic Summary


Distribution of 1 Blast Hits on the Query Sequence

Mouse over to see the details, click to show alignments

Color key for alignment scores

Score Range	Color
<40	Black
40-50	Blue
50-80	Green
80-200	Pink
>=200	Red

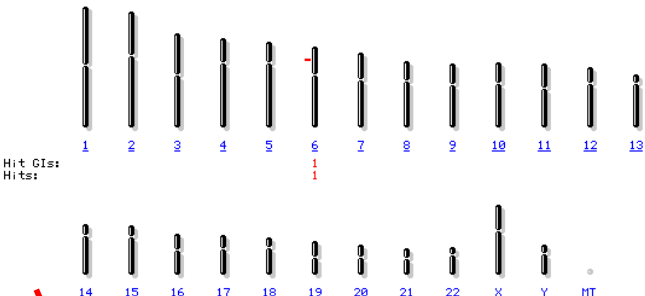
Query 0 50 100 150 200 250

NCBI  **NCBI Map Viewer**

PubMed Nucleotide Protein Genome Gene Structure

Search for on chromosome(s) assembly All

Homo sapiens (human) genome view
Build 37.1 statistics [Switch to previous build](#)



Hit GIs:
Hits:

Color key for scores: < 40 40-50 50-80 80-200 >= 200

BLAST search results: 1 BLAST hit found (Request ID "WC5RENPU01S").

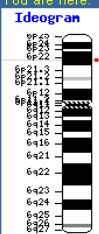
Chr	Map element	Type	BLAST results
			Hits Score E value
6	NT_007592	CONTIG	1 505 6e-141

36.1) Human genome overview page (Build 35.1)

[Map Viewer Home](#)

Map Viewer Help
Human Maps Help
FTP
Data As Table View
[Maps & Options](#)
Compress Map
Region Shown:

☐ out
☒ zoom
☐ in

You are here:


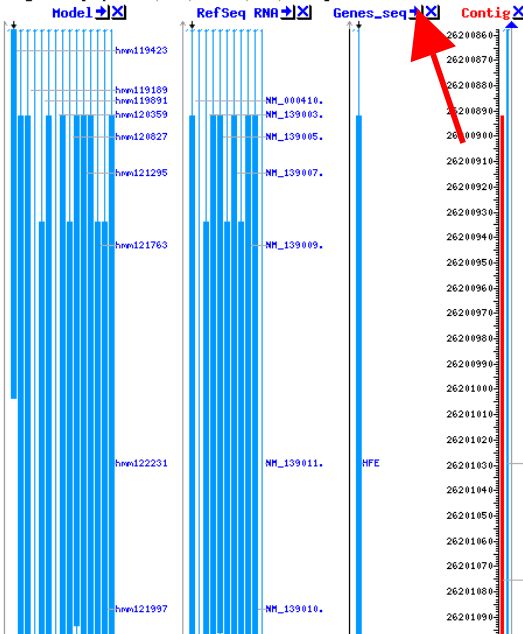
Query: [BLAST](#): (276 letters)

Color Key for Alignment Scores: <40 40-50 50-80 80-200 >=200

Master Map: Contig [Summary of Maps](#) [Maps & Options](#)

Region Displayed: 26,200,858-26,201,201 bp

[Model](#) [RefSeq](#) [RNA](#) [Genes_seq](#) [Contig](#) [Symbol](#) [Download/View Sequence/Evidence](#)



26200860
26200870
26200880
26200890
26200900
26200910
26200920
26200930
26200940
26200950
26200960
26200970
26200980
26200990
26201000
26201010
26201020
26201030
26201040
26201050
26201060
26201070
26201080
26201090
26201100

[NT_007592.14](#) ↓

[Blast hit](#) Identity=99% 1..276

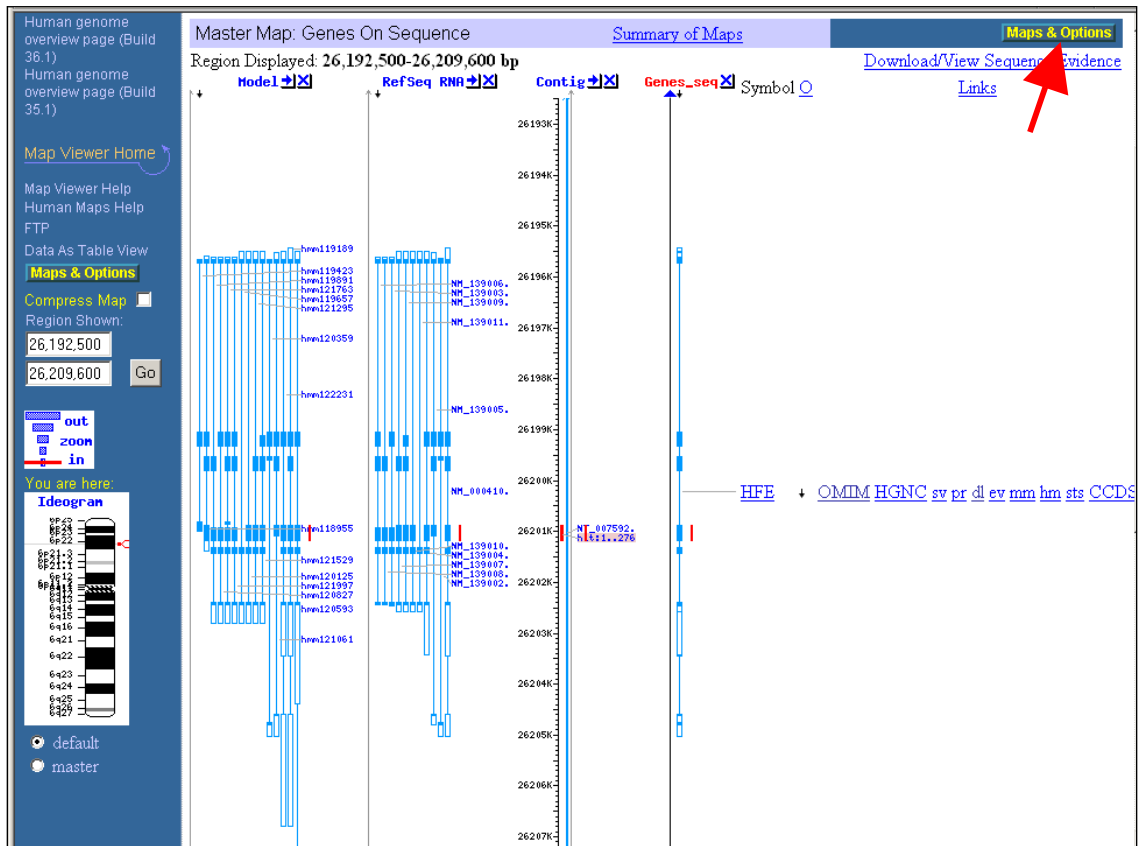
Range: from 16945699 to 16955310 [Show whole sequence](#) ☐ Reverse complemented strand [Refresh](#)

☐ 1: [NT_007592](#). Reports Homo sapiens chro...[gi:51465675]

```
>ref|NT_007592.14|Hs6_7749:16945699-16955310 Homo sapiens chromosome 6 genomic contig
GGGGACACTGGATCACCTAGTGTTCACAAAGCAGGTACCTTCTGCTGTAGGAGAGAGAGAACTAAAGTTC
TGAAAGACCTGTTGCTTTTACCCAGGAAGTTTACTGGGCATCTCCTGAGCCTAGGCAATAGCTGTAGGG
TGACTTCTGGAGCCATCCCGGTTTCCCCGCCGCCAAAAGAAAGCGGAGATTAAACGGGGACGTGCGGCCA
GAGCTGGGGAAATGGGCCCGGAGCCAGGCCGCGCTTCTCCTCTGATGCTTTTGACAGACCGCGGTCTCT
GCAGGGGCGCTTGCTGCGTGAGTCCGAGGGCTGCGGGCGAACTAGGGGCGCGGCGGGGTGGAAAAATCG
AAACTAGCTTTTTCTTTGCGCTTGGGAGTTTGCTAACTTTGGAGGACCTGCTCAACCCTATCCGCAAGCC
CCTCTCCCTACTTTCTGCGTCCAGACCCCGTGAGGGAGTGCCCTACCACTGAACTGCAGATAGGGGTCCCT
CGCCCCAGGACCTGCCCTTCCCCGGCTGTCCCGGCTCTGCGGAGTGACTTTTGGAAACCGCCCACTCCC
TTCCCCAACTAGAAATGCTTTTAAATAAATCTCGTAGTTCTCTCACTTGAGCTGAGCTAAGCCTGGGGCTC
CTTGAACCTGGAACCTCGGGTTTATTTCCAAATGTCAGCTGTGCAGTTTTTTCCCCAGTCATCTCCAAACAG
GAAAGTTCTTCCCTGAGTGCTTGCCGAGAAGGCTGAGCAAAACCCACAGCAGGATCCGCACGGGGTTTCCAC
CTCAGAACGAATGCGTTGGGCGGTGGGGGCGCGAAAGAGTGCGGTTGGGGATCTGAATTCTTCACCATTC
CAGCCACTTTTGGTGAGACCTGGGGTGGAGGTCTCTAGGGTGGGAGGCTCCTGAGAGAGGCCTACCTCGG
GCCTTTCCCCCACTCTTGGCAATTGTTCTTTTGCTTGGAAAATTAAGTATATGTTAGTTTTGAACGTTTGA
ACTGAACAATTCTTTTTCGGCTAGGCTTTATTGATTTGCAATGTGCTGTGTAATTAAGAGGCCTCTCTA
CAAAAGTACTGATAATGAACATGTAAGCAATGCACCTCACTTCTAAGTTACATTTCATATCTGATCTTATTG
ATTTTCACTAGGCATAGGGAGGTAGGAGCTAATAATACGTTTATTTTACTAGAAGTTAACTGGAATTCAG
ATTATATAACTCTTTTCAGGTTACAAAAGAACATAAAATAATCTGGTTTTCTGATGTTATTTCAGTACTAC
AGCTGCTTCTAATCTTAGTTGACAGTGATTTTGCCCTGTAGTGACACAGTGTCTGTGGGTCACACGC
CGGCCTCAGCACAGCACTTTGAGTTTGGTACTACGTGTATCCACATTTTACACATGACAAAGAAATGAGGC
ATGGCACGGCCTGCTTCTGGCAAAATTTATTCAAATGGTACACTGGGCTTTGGTGCCAGAGCTCATGTCTC
```

Result: The query EST is expressed by the HFE gene.

Step 3: Determine whether the EST contains any known SNPs



Organism: **Homo sapiens** [Help](#)

Chromosome: Region Shown:

Available Maps: **Maps Displayed (left to right):**

Org: Assembly:

([R] before map means 'ruler set')

More Options:

☐ Show Connections ☒ Verbose Mode

Compress Map: Auto Compress if > px

Page Length:

Thumbnail View: ☒ default (ideogram) ☐ master

Organism: **Homo sapiens** [Help](#)

Chromosome: Region Shown:

Available Maps: **Maps Displayed (left to right):**

Org: Assembly:

([R] before map means 'ruler set')

More Options:

☐ Show Connections ☒ Verbose Mode

Compress Map: Auto Compress if > px

Page Length:

Thumbnail View: ☒ default (ideogram) ☐ master

Organism: **homo sapiens** [help](#)

Chromosome: Region Shown:

Available Maps: **Maps Displayed (left to right):**

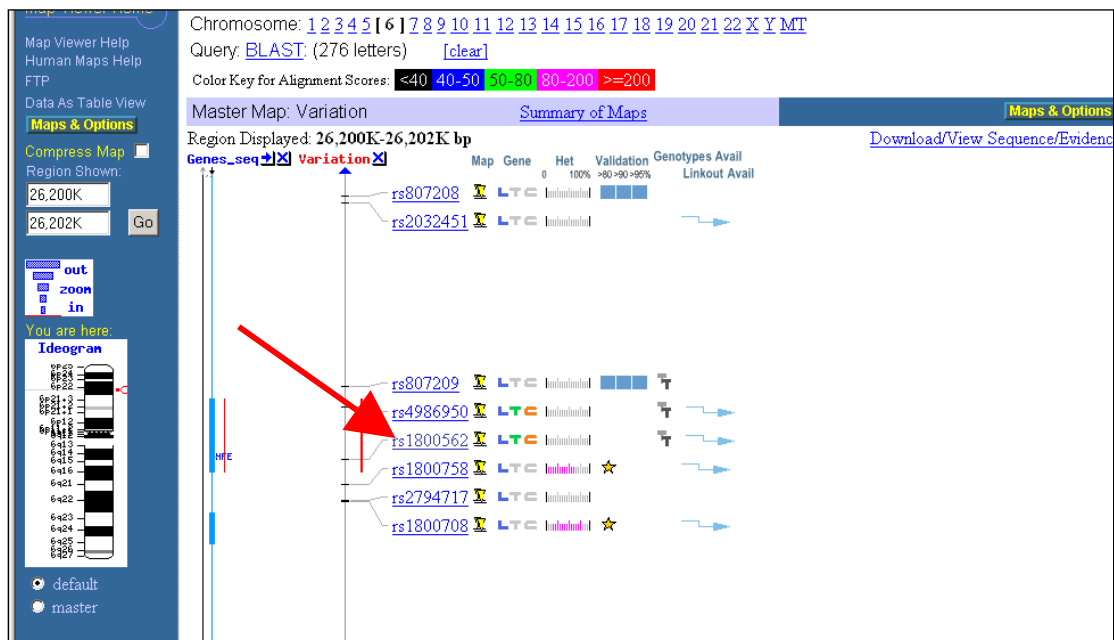
Org: Assembly:

-Sequence Maps-
 Ab initio
 Assembly
 BES_Clone
 Clone
 NCL_Clone
 Contig
 Component
 CpG Island

([R] before map means 'ruler set')

More Options:

☐ Show Connections ☒ Verbose Mode
 Compress Map: Auto Compress if > px
 Page Length:
 Thumbnail View: ☒ default (ideogram) ☐ master



Single Nucleotide Polymorphism

[PubMed](#)
[Nucleotide](#)
[Protein](#)
[Genome](#)
[Structure](#)
[PopSet](#)
[Taxonomy](#)
[OMIM](#)
[Books](#)
[SNP](#)

Search Entrez for

Reference SNP(refSNP) Cluster Report: rs1800562

refSNP ID: rs1800562 Organism: human (Homo sapiens) Molecule Type: Genomic Created/Updated in build: 89/126 Map to Genome Build: 36.2	Allele Variation Class: SNP: single nucleotide polymorphism Alleles: A/G Ancestral Allele: G	Links , Linkout
--	---	---

SEARCH

- Entrez SNP
- Blast SNP
- Batch Query
- By Submitter
- New Batches
- Method
- Population
- Detail
- Class
- Publication
- Chromosome Report
- Locus Information
- STS Markers
- Free Form Search

Fasta sequence (Legend)

>gn|dbSNP|rs1800562|allelePos=202|totalLen=450|taxid=9606|snpclass=1|alleles='A/G'|mol=Genomic|build=113

```

ATGTGAYCTC TTCAGTGACC ACTCTACGGT GTCGGGCCCTT GAACTACTAC CCCCAGAACA
TCACCATGAA GTGGCTGAAG GATAAGCAGC CAATGGATGC CAAGGAGTTC GAACCTAAAG
ACGTATTGCC CAATGGGGAT GGGACCTACC AGGGCTGGAT AACCTTGGCT GTACCCCTCG
GGGAAGAGCA GAGATATACG T
R
CCAGGTGGAG CACCCAGGCC TGGATCAGCC CCTCATTGTG ATCTGGGGTA TGTGACTGAT
GAGAGCCAGG AGCTGAGAAA ATCTATTGGG GGTTRAGAGG AGTGCCCTGAG GAGGTAAATTA
TGGCAGTGAG ATGAGGATCT GCTCTTTGTT AGGGGGTGGG CTGAGGGTGG CAATCAAAGG
CTTTAACTTG CTTTTTCTGT TTTAGAGCCC TCACCGTCTG GCACCCTAGT CATTGGAGTC
ATCAGTGG

```

SNP Details are organized in the following sections:

GeneView	Map	Submission	Fasta	Resource	Diversity	Validation
----------	-----	------------	-------	----------	-----------	------------

Integrated Maps (Hint: click on 'Chr Pos' or 'Contig Pos' column value to see variation in NCBI sequence viewer)

Genome Build	Chr	Chr Pos	Contig	Contig Pos	SNP to Chr	Contig allele	Contig to Chr	Group term	Group label	Contig label	Neighbor SNP	Map Method
37.1	6	26093141	NT_007592.15	26033141	+	G	+	GRCh37	GRCh37	GRCh37	view	blast
37.1	6	26093141	NW_260331.1	26033141	+	G	+	Celera	Celera	Celera	view	blast
37.1	6	26036201	NW_001838974.1	2047108	+	G	+	HuRef	HuRef	HuRef	view	blast
36.3	6	26201120	NT_007592.14	16951392	+	G	+	ref_assembly	reference	reference	view	blast
36.3	6	27322425	NW_922984.1	25717681	+	G	+	alt_assembly_1	Celera	Celera	view	blast
36.3	6	26036201	NW_001838974.1	2047108	+	G	+	alt_assembly_8	HuRef	HuRef	view	blast

GeneView

Organism: **Homo sapiens** [Help](#)

Chromosome: Region Shown:

Available Maps:

Org: Assembly:

([R.] before map means 'ruler set')

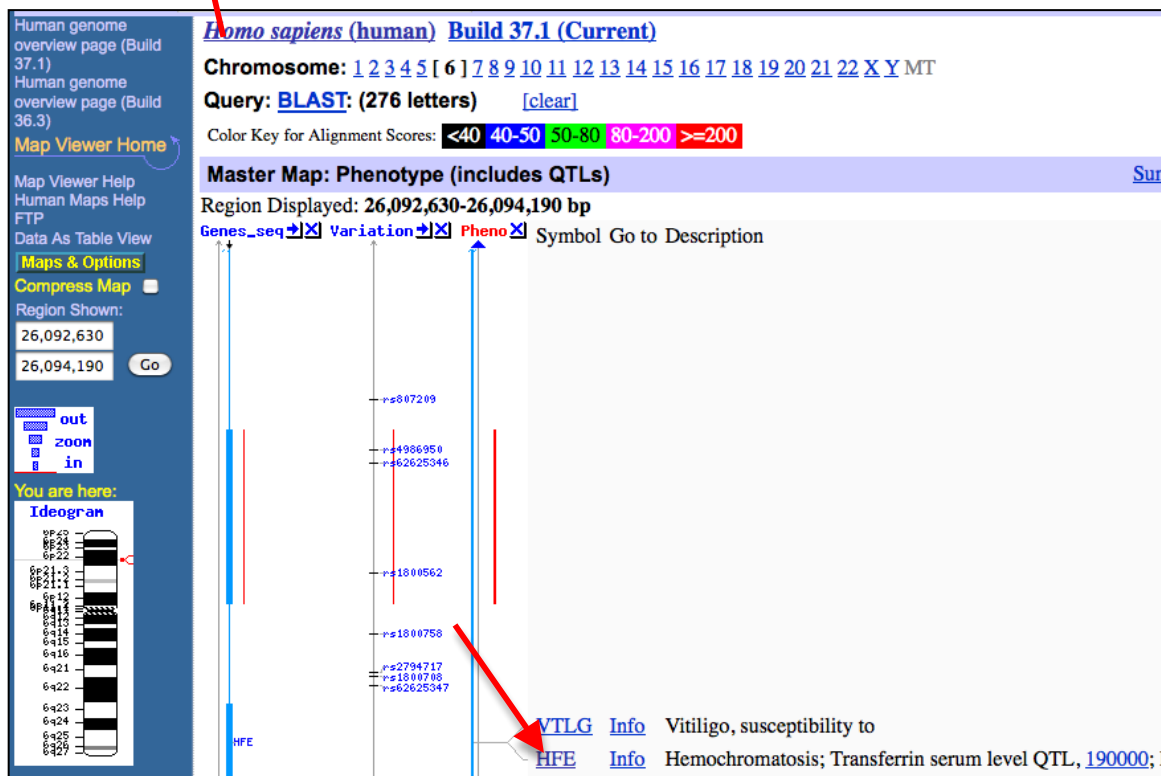
More Options:

☐ Show Connections
 ☒ Verbose Mode

Compress Map: Auto Compress if > px

Page Length:

Thumbnail View: ☒ default (ideogram) ☐ master



NCBI OMIM Online Mendelian Inheritance in Man Johns Hopkins University

My NCBI [Sign In] [Register]

All Databases PubMed Nucleotide Protein Genome Structure PMC OMIM

Search OMIM for [Go] [Clear]

Limits Preview/Index History Clipboard Details

Display Detailed Show 20 Send to

+235200 GeneTests, Links

HEMOCHROMATOSIS; HFE

Alternative titles; symbols

HLAH
HEMOCHROMATOSIS, HEREDITARY; HH
HFE GENE, INCLUDED; HFE, INCLUDED

Gene map locus [20p12, 6p21.3](#)

TEXT

DESCRIPTION

The clinical features of hemochromatosis include cirrhosis of the liver, diabetes, hypermelanotic pigmentation of the skin, and heart failure. Primary hepatocellular carcinoma (HCC; [114550](#)), complicating cirrhosis, is responsible for about one-third of deaths in affected homozygotes. Since hemochromatosis is a relatively easily treated disorder if diagnosed, this is a form of preventable cancer. 💡

At least 5 iron-overload disorders labeled hemochromatosis have been identified on the basis of clinical, biochemical, and genetic characteristics.

Classic hemochromatosis (HFE), an autosomal recessive disorder, is most often caused by mutation in a gene designated HFE on chromosome 6p21.3. It has also been found to be caused by mutation in the gene encoding hemojuvelin (HJV; [608374](#)), which maps to 1q21.

Entrez Gene
Nomenclature
RefSeq
GenBank
Protein

MIM +235200
Description
Clinical Features
Other Features
Inheritance
Mapping
Heterogeneity
Molecular Genetics
Genotype/Phenotype
Correlations
Diagnosis
Clinical Management
Population Genetics
Pathogenesis
Cloning
Biochemical Features
Gene Structure
Gene Function
Nomenclature
Animal Model
History
Allelic Variants
• View List
See Also
References
Contributors
Creation Date
Edit History

• Clinical Synopsis
• Gene map

NCBI OMIM Online Mendelian Inheritance in Man Johns Hopkins University

My NCBI [Sign In] [Register]

All Databases PubMed Nucleotide Protein Genome Structure PMC OMIM

Search OMIM for [Go] [Clear]

Limits Preview/Index History Clipboard Details

Display Allelic Variants Show 20 Send to

+235200

HEMOCHROMATOSIS; HFE

ALLELIC VARIANTS
(selected examples)

• **0001 HEMOCHROMATOSIS [HFE, CYS282TYR]**
PORPHYRIA CUTANEA TARDA, SUSCEPTIBILITY TO, INCLUDED
PORPHYRIA VARIEGATA, SUSCEPTIBILITY TO, INCLUDED
HEMOCHROMATOSIS, JUVENILE, DIGENIC, INCLUDED
ALZHEIMER DISEASE, SUSCEPTIBILITY TO, INCLUDED
TRANSFERRIN SERUM LEVEL QUANTITATIVE TRAIT LOCUS, INCLUDED
MICROVASCULAR COMPLICATIONS OF DIABETES, SUSCEPTIBILITY TO, 7, INCLUDED

• **0002 HEMOCHROMATOSIS [HFE, HIS63ASP]**
MICROVASCULAR COMPLICATIONS OF DIABETES, SUSCEPTIBILITY TO, 7, INCLUDED

• **0003 HEMOCHROMATOSIS [HFE, SER65CYS] dbSNP**

• **0004 HFE INTRONIC POLYMORPHISM [HFE, 5569G-A]**

• **0005 HFE POLYMORPHISM [HFE, VAL53MET] dbSNP**

• **0006 HFE POLYMORPHISM [HFE, VAL59MET] dbSNP**

• **0007 HEMOCHROMATOSIS [HFE, GLN127HIS] dbSNP**

• **0008 HEMOCHROMATOSIS [HFE, ARG282MET]**

Entrez Gene
Nomenclature
RefSeq
GenBank
Protein

MIM +235200
Description
Clinical Features
Other Features
Inheritance
Mapping
Heterogeneity
Molecular Genetics
Genotype/Phenotype
Correlations
Diagnosis
Clinical Management
Population Genetics
Pathogenesis
Cloning
Biochemical Features
Gene Structure
Gene Function
Nomenclature
Animal Model
History
Allelic Variants
• View List
See Also
References
Contributors
Creation Date
Edit History

• Clinical Synopsis
• Gene map

Result: Mutations in the HFE gene are associated with hemochromatosis disease

Problem 2:

A laboratory has generated an EST library from a sickle cell anemia patient and wants to identify the gene(s) causing the phenotype. Sickle cell anemia is a disease in which the red blood cells are curved in shape, and which causes pain and fever.

Outline:

We will follow these steps to solve the problem:

1. Compare an EST from a sickle cell anemia patient to the human genome (using BLAST).
2. Identify the gene(s) aligning with the EST and download their sequences (using Map Viewer).
3. Identify whether the EST contains any known nucleotide variations (single nucleotide polymorphisms) (using dbSNP).
4. Determine whether a mutant form of the gene is known to cause a phenotype (using OMIM).

Step 1. Compare ESTs to the human genome (using BLAST):

One way to identify the genes expressing the ESTs is to compare the EST sequence with the human genome assembly and the genes annotated on it. To access the specialized BLAST page for searching against the human genome assembly, click on

BLAST (human genome)

Copy the EST sequence provided below. Paste it in the query box of the BLAST page and select the “genome (reference only)” database. Start the search by clicking on the “Begin Search” button.

```
ACATTTGCTTCTGACACAACCTGTGTTCACTAGCAACCTCAAACAGACACCAT  
GGTGCATCTGACTCCTGTGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAA  
GGTGAAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAGGCTGCTGGTGG  
TCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCC  
TGATGCTGTTATGGGCAACCCTAAGGT
```

Name the chromosome and the contig that we get as a BLAST hit. Note that the similarity is on the minus strand of genome. Is the EST sequence 100% identical to the genomic sequence? Note the nucleotide difference and position between the two sequences.

Next, open another window and use Align 2 or more sequences (linked on the BLAST homepage) and align the EST sequence to NC_000011, the chromosome record for chromosome 11. The results will appear similar, but note the position of the mismatch relative to the chromosome. The coordinates are slightly different between the contig and chromosome records. You will need the chromosome coordinates for the next part of the exercise. Now return to the previous window where you ran the human genome BLAST query.

Step 2. Identify the gene(s) expressing the EST and download their sequences:

To visualize the BLAST hit on the genome using Map Viewer, click on the "Genome View" button at the top of the results page, then on the Map element "NT_009237". Currently, 4 maps should be displayed (Contig, Model, RNA and Gene_seq). Zoom out 2 or 4 times by clicking on right most contig map and selecting the appropriate option.

The best BLAST hits, indicated by the red bars, are in the region of two exons of the HBB gene annotated on the human genome. Make the Gene_seq map a master map by clicking on the arrow at the top of the map. Note that the gene is annotated on the minus strand. To display the entire HBB gene sequence, click on the "dl" link, choose minus strand from the pull down menu, click on "Change Region/Strand" and display the sequence by clicking on "Display". Copy the sequence and paste it in the area provided below. We will use it later to obtain the exon-intron structure. You can adjust the nucleotide locations to download the upstream or downstream sequence by using the "adjust by" and "Change Region/Strand" option.

Step 3. Determine whether the EST contains known SNPs:

Go back to the Map Viewer report. Click on the Maps and Options link. Remove all the maps except the Gene_seq map by selecting the map under the Maps Displayed menu and clicking on Remove. Now add the variation map from the Available maps menu (by selecting the map and clicking on Add). Make the Variation map as the master map by selecting it and clicking the Make Master/Move to Bottom option. Then click on Apply. Now two maps are displayed, Variation (it's the rightmost and the master map) and Gene_seq. The master map provides detailed information for the map features, in this case SNPs. (The Map Viewer exercise describes the usage of the Map Viewer in detail.) Zoom in on the blast hit area (red bar). There are numerous SNPs in the area, one of them is rs334. To see it, type in the exact genomic coordinates in the region shown box (5,248,232) Click on the link for the SNP. There is an A/T SNP is at the nucleotide position 5188232 on the contig NT_009237 as mentioned under Fasta sequence and Integrated maps. Is this the same nucleotide variation found in the BLAST result in Step 1?

Step 4. Determine whether the mutant HBB gene causes a phenotype:

Go back to the Map Viewer report. Go back to the Map Viewer report and add the Phenotype map as the master map using the Maps and Options menu. Click on the HBB link that leads to the hemoglobin beta record (141900) in the OMIM database. The record details how mutations in the HBB gene are associated with a phenotype, sickle cell anemia. As mentioned in the report, the allelic variants are listed for the mature HBB protein which lacks initiator methionine. Click on the Allelic Variant “View list” to get information about mutant proteins from patients. Is Glu6Val variant mentioned in the list? Which phenotype does it cause?

Summary:

This exercise describes steps to identify the gene expressing the ESTs obtained from a sickle cell anemia patient, download the gene sequence, identify known SNPs in the gene and find SNP-associated phenotypes.

Step 1: The query EST sequence was found to align to contig NT_009237.17 on chromosome 11 with one nucleotide difference (T to A with respect to the nucleotide 5188232 on the contig).

Step 2: The query EST was found to be expressed by the HBB gene.

Step 3: The query EST sequence contains a known SNP (T/A with respect to the nucleotide 5188232 on contig NT_009237.17).

Step 4: Mutations in the HBB gene are associated with sickle cell anemia.